An Efficient and Cost-Effective Protocol for Selecting Transcription Factor Binding Sites that Reduces Isotope Usage

*Joseph Krystel and Kasirajan Ayyanathan**

Center for Molecular Biology and Biotechnology, Florida Atlantic University, Jupiter, Florida 33458, USA

To function, transcription factors must position themselves by binding to DNA in a sequence-specific manner. Knowing the binding sites of these factors is a necessary step in understanding their activity. The standard protocols used for selecting a consensus-binding sequence for a DNA binding domain often require the use of radioisotopes to attain the necessary level of power in the assay. Alternatives are often less sensitive and may require an expensive apparatus for visualizing. We have created a modified binding site selection (BSS) protocol to improve efficiency and decrease the use of radioisotope. A GST affinity-tagged DNA binding domain construct was immobilized on a GSH affinity column and used to select from a randomized oligonucleotide library identical to those typically used in a radiolabeled BSS protocol. This produced a library specifically pre-enriched for use in a standard sequential EMSA selection. Use of a pre-enriched library reduced the total number of labeled rounds required for selection, decreasing the use of radioisotope while maintaining efficacy. The protocol was used to select for the binding sequence for several *Drosophila melanogaster* transcription factors. The consensus sequence was then shown by competitive binding experiments to associate with the protein in a sequence-dependent manner.

KEY WORDS: EMSA, radioactivity, affinity chromatography

INTRODUCTION

One truth revealed in our recent understanding of genetics is the importance of gene regulation. The number or length of genes in an organism does not directly correlate to the complexity of structures or behaviors. More important is the ability to regulate precisely when and how those genes are expressed. Of the many participants in this control, transcription factors are among the best characterized. Their activity is based on interaction with specific DNA sequences that then position the machinery needed to modify the local chromatin structure into a conformation, which excludes or facilitates binding of the transcription preinitiation complex. One of the most common types of transcription factor in humans and all other eukaryotes is the C_2H_2 zinc finger protein transcription factor (ZFP).[1] First identified in the late 1980s, these proteins are so named for their DNA binding domain, which consists of a tandem array of C_2H_2 zinc finger domains.[2] They are further grouped based on the presence of one of several different possible effector domains that provide the regulatory activity. Possible effector domains and their associated families include the Krüppel-associated box (KRAB) domain, BR-C, ttk, and BTB/Pox virus and zinc finger domain, zinc finger-associated (ZAD) domain, and SCAN domain.[3] Understanding these transcription factors is key to understanding how complex organisms develop. Discovery, structure, and biomedical applications are reviewed in a recent article by Klug.[4]

In addition to the implications inherent in understanding the control systems that regulate biological development, further understanding of the molecular functions of these transcription factors has biomedical and research potential. In recent years, several groups have been working on ways to adapt the very versatile ZFP system to design customized transcription factors for specific applications. As the DNA binding domain acts independently of the effector domain and as the individual zinc finger domains can be assembled into an array to target a larger, less common sequence, it is possible to target a domain containing any desirable activity to nearly any sequence properly positioned to regulate a gene of interest.[5,6] That activity may be in the form of a transcriptional repressor or activator or even nuclease activity as a precursor for using homologous recombination.[7–9]

Identifying the specific nucleotide sequence bound by a transcription factor is an important early step in charac-

*ADDRESS CORRESPONDENCE TO: Kasirajan Ayyanathan, Center for Molecular Biology and Biotechnology, Florida Atlantic University, 5353 Parkside Dr., MC-17, RF 203, Jupiter, FL 33458, USA (Phone: 561-799-8322; Fax: 561-799-8332; E-mail: kayyanat@fau.edu).
doi: 10.7171/jbt.12-2302-004

terizing its molecular and biological functions. This may be carried out by selecting for the DNA sequence that most efficiently binds the protein or by locating the native targets within the genome. Both approaches have advantages and disadvantages.

Techniques to identify sequences that efficiently bind a transcription factor may be constructed in different ways, but in general, they involve repeated selections of a random collection of oligonucleotides against the protein of interest. As only a very small number of the total sequences present in a random oligonucleotide library will bind the protein, this method requires multiple rounds of selection and a very powerful examination technique to identify the signal. The archetypal method for this approach used double-stranded oligonucleotides radiolabeled with ^{32}P. The oligonucleotides each contain a variable region large enough to represent the expected binding site and short enough to make analysis reasonably possible. Zinc finger arrays generally recognize three or four nucleotides/finger participating in the binding.[10,11] The labeled family of oligonucleotides is then bound under near-cellular conditions to a purified form of the protein. That protein may be obtained by way of antibody selection or generated in an ex vivo expression system and then purified by affinity chromatography. Once bound, the combined protein and DNA sample is then run on a nondenaturing polyacrylamide gel. Oligonucleotides, which are unbound, will pass quickly to the end of the gel, whereas those incorporated into a larger DNA-protein complex will be retarded in their mobility. This shift separates the two populations of molecules. Those that bind the protein may then be recovered from the gel and amplified by PCR into a new library, enriched in molecules that bound the protein. The binding is then repeated multiple times to further enrich the library, until a point when it contains primarily those members that efficiently bind the protein. That library is then sequenced, and an analysis of the sequences for shared motifs will identify a consensus-binding sequence. Several protocols, involving immobilized protein on filters or Southwestern blot analysis, have also been reported.[12,13] More recent adaptations of this methodology have been developed, which incorporate nonisotopic methods of isolating the DNA-protein complex from free DNA and high-throughput sequencing methods to increase the resolution of the binding sequence.

Identifying a binding site consensus by this method does not directly identify genes within the organism that are bound and regulated by the protein. Other in vivo conditions, such as the presence or absence of cofactors, the binding of other more strongly associated transcription factors, and epigenetic factors, may all play significant roles in this process. It will, however, provide a means to identify potential target genes, even those that may not be associated with the transcription factor under the growth conditions typically used. It also provides information about the molecular function of the zinc fingers in the array. As zinc finger arrays are modular in function, this information can be invaluable in constructing artificial transcription factors for genetic research.[14]

The second approach for identifying binding sequences for a transcription factor is to collect and isolate those sequences bound in vivo by the protein. A common protocol would involve binding the protein to a sample of naked genomic DNA and then exposing the bound DNA to a very low concentration of DNase. Those portions bound by the protein will be shielded from the nuclease activity and will not be degraded. Affinity chromatography or antibody precipitation can be used to isolate these fragments of DNA, which may then be sequenced.[15] This method does immediately provide the researcher with regions targeted by the transcription factor under the conditions used. However, the sequence may also include nucleotides unnecessary to the binding activity. Bases that are physically covered by the protein may not be involved directly in the protein-DNA interaction. Those sequences may be maintained in the genome by positive selection as a result of targeting by another transcription factor with a binding site that overlaps for regulatory reasons. This has been shown to be possible in genes, such as muscle regulatory factor 4, where overlapping binding sites for TATA-box binding protein and myocyte enhancer factor-2 are contained in one region (-26 to -15), and specific nucleotides are required for each binding.[16] They may also be conserved as binding sites of another protein that acts as a binding site competitor for regulation of transcription factor activity. This is seen in the binding of enhancer factor 1 to the E2 box (G/ACAGNTGT/G).[17]

Other more gene-specific methods are also possible. If a known target gene has been identified through other means, a sequential deletion of its upstream regulatory region can reveal the region necessary for the binding activity; if then, combined with EMSA against oligonucleotides tailored to mimic that region, a single, sufficient binding sequence may be characterized. An example of this method can be found in Harms et al.[18]

An additional strategy has been used, beginning with a known DNA sequence, and presenting to it, by means such as phage display, a wide array of different C_2H_2 zinc finger domains.[19] This method does not provide any direct information about a particular transcription factor, but it does allow for the creation of zinc finger libraries to construct artificial transcription factors. The method is complicated by the contextual effect, wherein adjacent zinc fingers slightly affect the binding activity, so multiple rounds of

design may be required before the desired binding is achieved.[20,21] Therefore, the conventional method of binding site selection (BSS), which involves multiple rounds of selection from a radiolabeled oligonucleotide library, still remains the most common choice. However, in this article, we describe a variation of the conventional protocol for selecting the transcription factor binding sites that is simple, robust, and uses much lower amounts of radioactive isotopes.

MATERIALS AND METHODS

GST-ZFP Fusion Protein Production

We developed and used this modified BSS protocol to investigate several members of the ZAD family of transcription factors in *Drosophila melanogaster*, including the protein encoded from CG4413 (Trade Embargo). ZAD is the largest family of C_2H_2 transcription factors in *Drosophila* and is analogous to the KRAB family in mammals.[22] We amplified the sequences for the DNA binding domains from multiple ZAD members, including those contained within residues 272–406 of the *Drosophila* gene CG4413 from a 0- to 8-h embryonic *Drosophila* cDNA library using 5′ *BamHI*-gtgGTCGACtgcgtcatttctccaggtg and 3′ *SalI*-gtgGGATCCggcagaaagccgcgcaacaag primers, which were designed to contain built-in restriction sites (underlined) for *BamHI* and *SalI*. The resulting product was then ligated in-frame with the pGEX 4T-2 expression plasmid, which contributed a GST affinity tag and inducible control by addition of IPTG. Multiple independent clones were produced and transformed into DH5α *Escherichia coli* cells for increased plasmid production.

Each independent clone was then cultured for mini-plasmid DNA preps. The plasmids were then checked for the correct insert size by restriction endonuclease digestion and gel electrophoresis. Confirmed recombinant plasmids from each pGEX-ZAD construction were then transformed into the BL21 *E. coli* host for protein expression. Transformed BL21 cells were cultured in lysogeny broth/ampicillin/kanamycin media and tested via overnight IPTG (1 mM) induction at 37°C for protein production. A noninsert-bearing pGEX-4T2 plasmid containing culture was used as a control to check for the proper size protein production. Two expression clones were then selected from each construction for maxi-protein production by inducing a 500-ml culture with 0.1 mM IPTG at 30°C. Expressed proteins were released from the cells by lysozyme treatment, followed by sonic disruption. The soluble fractions of each protein, predicted to represent the functional form, were then bound on a GSH bead column, eluted in a 15-mM reduced GSH containing Tris-buffered elution buffer, and dialyzed later to remove the reduced GSH.

DNA BSS

Proteins were bound in 1.5 ml microcentrifuge tubes containing 10 μl GSH beads, supplemented with 10 μl G75 sepharose beads to increase the volume for rinses. The unbound proteins were removed with three 800 μl washes of PBS wash buffer (137 mM NaCl, 2.7 mM KCl, 10 mM sodium phosphate dibasic, 2 mM potassium phosphate monobasic, 1 mM PMSF, 0.5% BSA, 0.5 mM DTT). Samples were rotated at 25°C for 2 min and centrifuged at 2000 RPM for 5 min, and the supernatants were removed by vacuum. The samples were then rotated at 4°C for at least 30 min in PBS wash buffer to block nonspecific binding. Members of a 49mer oligonucleotide library were then bound to the immobilized protein in 1× 20 mM HEPES, 75 mM NaCl, 0.5 mM DTT, 10% glycerol, 0.5 mM $MgCl_2$, and 50 μM $ZnSO_4$ (NEBB) wash buffer (1 mM PMSF, 0.5% BSA, 0.5 mM DTT), rotated for 30 min at 4°C, followed by 30 min at 25°C. The library consisted of annealed pairs of oligonucleotides of the species 5′-agacGGATCCattgca-NNNNNNNNNNNNNN-NNNN-ctgtccGAATTCgga-3′; each member contained a random 18N central region, which was flanked by known primer targets with embedded *BamHI* and *EcoRI* restriction sites (restriction sites underlined). DNA-protein complexes were eluted from the beads in 20 μl 15 mM reduced GSH containing Tris elution buffer. Selected oligonucleotides present in the elution were then amplified by the known primer targets to create a library enriched in the sequences that can effectively bind the protein construct. This enriched library was used in a second round of enrichment, with the products then used for a third and fourth round. The fourth round library was then taken for two more rounds of labeled BSS.

The labeled rounds of selection were conducted as described.[23] Each protein was combined with a ^{32}P-ATP end-labeled 49mer oligonucleotide library created from the enriched library discussed above. The protein-DNA binding was conducted in nuclear extract binding buffer (NEBB) and ran on a 5% polyacrylamide gel for EMSA. The GST-zinc finger (ZnF)-oligonucleotide complexes were electro-eluted from the gel and amplified by PCR using the known flanking sequence primers. The resulting products were run on a 10% PAGE gel and stained in a solution of 0.0125% ethidium bromide for 20 min.

The PCR products were prepared for use by proteinase-K treatment, phenyl-chloroform-isoamyl alcohol extraction, and chloroform-isoamyl extraction, followed by ethanol precipitation. These enriched libraries were then used in the second round of mobility shift assays with each of the purified GST-ZnF proteins. Bound oligonucleotides from the second enrichment were amplified by PCR and digested with *EcoRI* and *BamHI* restriction enzymes pres-

ent in the flanking region of each oligonucleotide. These digested products were then ligated into pUC18 plasmid vector for cloning and transformed into DH5α cells. Multiple, independent clones were produced. Mini-plasmid preps were conducted using each clone. The plasmid DNA was checked for the presence of an insert by way of enzyme digestion. The resulting positive clones for each construct were then sent to the Interdisciplinary Center for Biotechnology Research Genomics Core (University of Florida, Gainesville, FL, USA) for high-throughput cycle sequencing in a 96-well format.

RESULTS

The modified, robust BSS protocol, which only uses a minimal amount of radioactive isotope is represented diagrammatically in Fig. 1. As seen in the protocol, the enriched library obtained after four rounds of unlabeled selection was radiolabeled and an EMSA performed. A 4-h exposure of the second selection round for three ZFP constructs, including CG4413, can be seen in Fig. 2A. Original pre-enriched and selected libraries were counter-selected against a GST control, and as expected, no complexes were observed (data not shown). After the second round of labeled selection and EMSA, the bound DNA molecules were released from the DNA-protein complexes by electroelution and taken for cloning. At least 20 insert sequences were used in building a binding site consensus

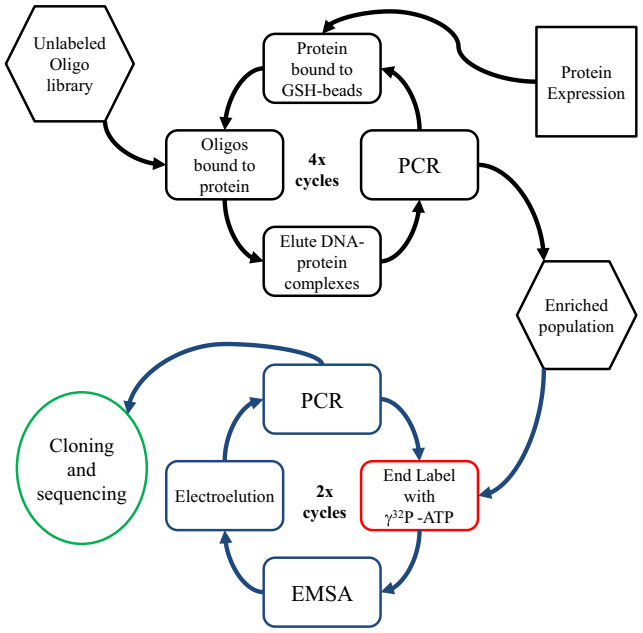

FIGURE 1

A schematic representation of the coupled cold and hot binding protocol. A random dsDNA library is selected first against proteins immobilized on GSH affinity beads and then in a traditional gel shift assay.

for each of the selected ZAD members. For CG4413, 29 insert sequences were identified from a total of 25 independent clones. Two clones contained multiple inserts ligated head-to-tail into a single plasmid vector. A binding site consensus sequence was derived that substantially matched 18 of the insert variable regions. This consensus was derived by use of the publically available ClustalW tool from the European Bioinformatics Institute. We began by aligning all of the 18N variable sequences from all of the independent clones. Conserved regions found in this analysis were then repeatedly aligned in batches that excluded dissimilar sequences to minimize the effect of outliers on the nature of the sequence. We expected a degree of plasticity in the final binding sequence and the possible incorporation of less than ideal members in the BSS, based on previously published work in the field.[7] We therefore selected the conserved region incorporating the most independent members as a consensus sequence. Nucleotide positions located in the invariant flanking regions of the oligonucleotide or with low quality sequencing data were excluded from the consensus building. This consensus was of the type 5′-CCCCTTGCCCYYCCCC-3′. A summary of the sequences used to derive this consensus is shown in Table 1.

This consensus sequence was then tested for its ability to bind the GST-ZnF protein used in the selection and for the specificity of that binding. This testing included a competition binding with two double-stranded oligonucleotides. The first contained the consensus binding site sequence (WT), and the second was a scrambled (SC) oligonucleotide designed to avoid duplicating portions of the consensus. We produced the oligonucleotides 5′-GCGCCCCTTGCCCCTCCCC-3′ (WT) with its reverse complement and 5′-TATACTATAGCATATATAC-3′ (SC) with its reverse complement and annealed each into double-stranded forms. The forward strands of each sequence used for competition are also represented in Table 1. One population of the WT oligonucleotide was end-labeled by DNA kinase with γ-^{32}P ATP and bound to the purified CG4413 DNA binding domain fusion protein. The oligonucleotide derived from the consensus sequence did produce a shifted band that was not seen in the absence of protein or labeled oligonucleotide, indicating that the sequence is bound effectively by the protein. This binding was then subjected to increasing concentrations of the selected (WT) or SC oligonucleotides to test their ability to displace the bound and labeled oligonucleotide. The WT oligonucleotide was significantly more effective at displacing the complex, indicating that the binding is dependent on the specific sequence. The competition gel autoradiogram is shown in Fig. 2B.

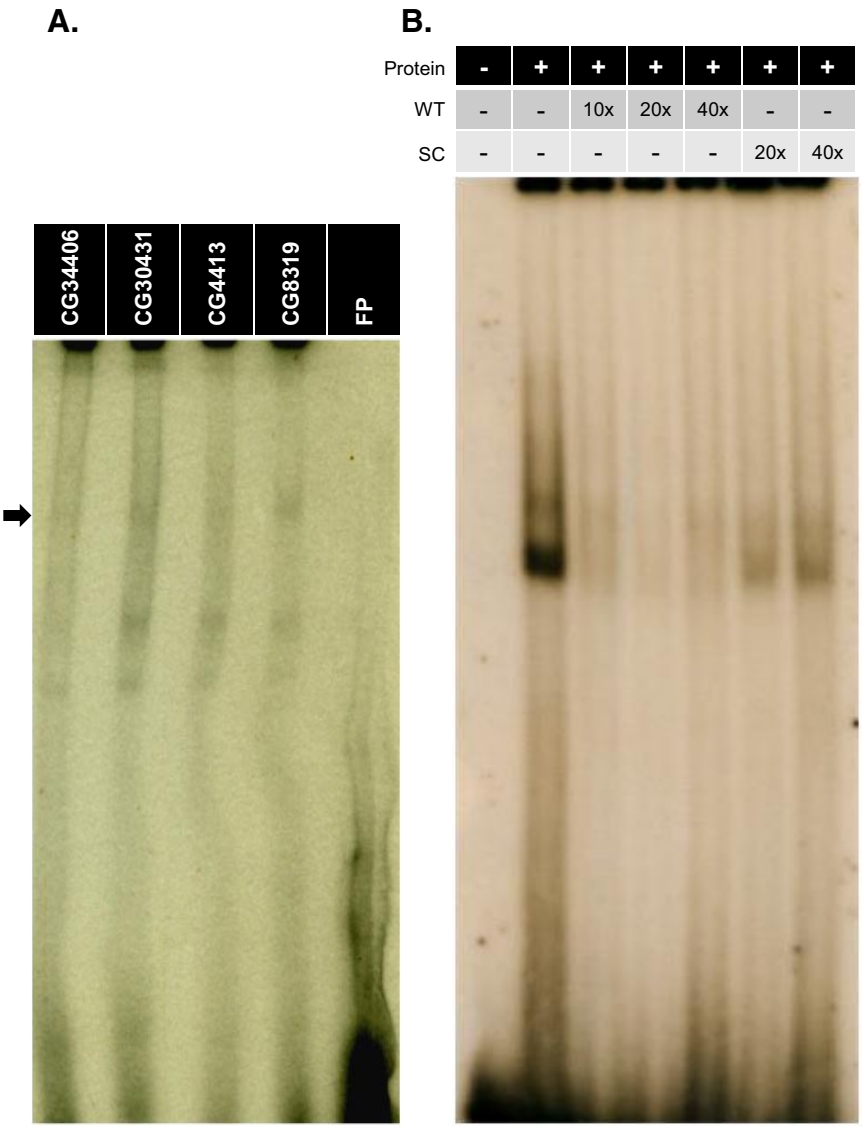

FIGURE 2

(A) BSS and competitions. EMSA showing the second round of labeled selection for four GST-ZnF construct proteins, CG34406,[1] CG30431,[2] CG4413,[3] CG8319,[4] and the labeled oligonucleotide library with a free probe control (FP). Binding reactions were run on a compact, 5% native DNA-PAGE. The position of the complex eluted from the gel is marked by a black arrow. (B) The final selected consensus sequence was end-labeled and added to each lane. Each protein (+) lane was allowed to bind to purified DNA binding domain/GST fusion protein. Competition was conducted using unlabeled consensus (WT) or scrambled (SC) oligonucleotides of increasing concentration, from 10× the labeled oligonucleotide concentration to 40× the labeled oligonucleotide. Competition binding reactions were run on a longer, 4% native DNA-PAGE.

DISCUSSION

This modified technique is particularly well suited for adoption in labs that are currently using traditional radiolabeled BSS protocols. The similarities in the pre-enrichment rounds to the standard EMSA enrichment allow for an easy transition. Affinity-tagged fusion proteins and their matching affinity beads are already widely used in these studies as a means of making and purifying large quantities of the proteins and do not need to be purchased solely for the pre-enrichment purposes. Eluted fractions from these cold bindings were PCR-amplified using the same primers and conditions already required for the EMSA enrichment rounds. The only additional equipment required was an inexpensive vacuum pump for removing supernatant from the binding reactions.

Many alternative techniques for reducing or fully eliminating the use of radioisotope have been described in the literature. The proliferation of these systems is a testament to the desirability of decreasing radiation work. Each of these alternatives must in some way replace the powerful, selective function of the radioisotope. This may be done by replacing the radioisotopes for visualizing the EMSA with a fluorescent tag or colorimetric label. These protocols require the incorporation of related tags into the complex partners. This is a step that requires an additional step, subject to the limitations of the process, including the inefficiency of related enzymes, such as TdT, and yields a product that may physically interfere with the complex formation. The alternative methods may also use an entirely different complex separation technique for complex separation, such as immobilization of the protein. This is similar to our pre-enrichment procedure, but without the coupled isotopic rounds of selection, it must be paired with a high-throughput sequencing and computational analysis, such as is reported in Reiss and Mobley.[24] The cost of these systems put them beyond the reach of many labs. It is also

TABLE 1

Binding Consensus Shown for Clones 1[en]18 and Quantified in Number of Matches and Percentage Present

			C	**C**	**C**	**C**	**T**	**T**	**G**	**C**	**C**	**G**	**G**	**T**	**C**	**C**	**C**	**C**
			C	**C**	**C**	**C**	**T**	**T**	**G**	**C**	**C**	**G**	**G**	**T**	**C**	**C**	**C**	**C**
			C	**G**	**C**	**C**	**T**	**T**	**G**	**C**	**C**	**C**	**C**	**T**	**C**	**C**	**C**	**C**
			C	**G**	**C**	**C**	**T**	**T**	**G**	**C**	**C**	**C**	**C**	**T**	**C**	**C**	**C**	**C**
			G	**C**	**C**	**A**	**C**	**T**	**G**	**C**	**C**	**G**	*	*	*	*	*	*
			C	**C**	**C**	**G**	**G**	**T**	**G**	**C**	**C**	**C**	**T**	**A**	**G**	*	*	*
			C	**T**	**C**	**C**	**T**	**T**	**G**	**C**	**C**	**G**	**C**	**G**	**C**	**G**	**C**	**G**
			C	**G**	**C**	**G**	**C**	**C**	**T**	**G**	**C**	**C**	**C**	**C**	**G**	*	*	*
			T	**G**	**C**	**T**	**C**	**T**	**T**	**C**	**C**	**C**	**A**	**C**	**C**	**C**	**G**	**G**
			G	**G**	**T**	**C**	**A**	**G**	**T**	**C**	**C**	**G**	**T**	**T**	**G**	*	*	*
			*	**C**	**T**	**G**	**G**	**T**	**G**	**C**	**T**	**G**	**C**	**C**	**C**	**T**	**A**	**C**
			*	**T**	**G**	**C**	**G**	**T**	**G**	**C**	**T**	**C**	**A**	**G**	**T**	**G**	**C**	**T**
			*	**T**	**G**	**C**	**G**	**T**	**G**	**C**	**T**	**C**	**A**	**G**	**T**	**G**	**C**	**T**
			*	**C**	**T**	**C**	**T**	**T**	**G**	**C**	**T**	**C**	**T**	**G**	**G**	**T**	**A**	**T**
			G	**G**	**G**	**C**	**C**	**T**	**C**	**C**	**C**	**G**	**T**	**G**	**T**	**G**	**G**	**G**
			A	**C**	**C**	**A**	**T**	**T**	**C**	**C**	**C**	**C**	**A**	**A**	*	*	*	*
			*	*	*	*	**C**	**T**	**G**	**C**	**C**	**C**	**C**	**C**	**T**	**C**	**C**	**C**
			C	**C**	**C**	**G**	**G**	**T**	**G**	**C**	**C**	**C**	**T**	**A**	**G**	*	*	*
			13	**17**	**17**	**16**	**18**	**18**	**18**	**18**	**18**	**18**	**17**	**17**	**16**	**12**	**12**	**12**
		A	1	0	0	1	2	0	0	0	0	0	4	3	0	0	2	0
		T	1	3	3	1	**7**	**16**	3	0	4	0	**5**	**5**	4	2	0	3
		C	**8**	**8**	**11**	**10**	4	1	2	**17**	**14**	**11**	**6**	**4**	**7**	**6**	**8**	**6**
		G	3	6	3	4	5	1	**13**	1	0	7	2	5	5	4	2	3
		A	8	0	0	6	11	0	0	0	0	0	24	18	0	0	17	0
		T	8	18	18	6	**39**	**89**	17	0	22	0	**29**	**29**	25	17	0	25
		C	**62**	**47**	**65**	**63**	22	6	11	**94**	**78**	**61**	**35**	**24**	**44**	**50**	**67**	**50**
		G	23	35	18	25	28	6	**72**	6	0	39	12	29	31	33	17	25
			C	**C**	**C**	**C**	**T**	**T**	**G**	**C**	**C**	**C**	**Y**	**Y**	**C**	**C**	**C**	**C**
		WT oligo																
G	**C**	**G**	**C**	**C**	**C**	**C**	**T**	**T**	**G**	**C**	**C**	**C**	**C**	**T**	**C**	**C**	**C**	**C**
		Mutant oligo																
T	**A**	**T**	**A**	**C**	**T**	**A**	**T**	**A**	**G**	**C**	**A**	**T**	**A**	**T**	**A**	**T**	**A**	**C**

*, Empty position regions of low-quality sequencing or the end of the oligonucleotide variable region. Bolded values represent those nucleotide populations selected for building the consensus.

possible to wholly remove the need for complex-FP separation by performing each potential binding independently on microarray, such as is reviewed in Wang et al.[25] This method is limited by the overall length of the sequence used in each reaction and requires technology and apparatus, not necessarily available at all institutions. In contrast to these methods, our protocol requires no significant apparatuses or materials not already available in a lab equipped for isotopic BSS. It also requires no additional skills or training beyond those already used in the preexisting methods. For essentially no cost in terms of funding, time, or training, a lab may transition from the traditional methods to this modified protocol and reduce the overall radiation use.

This protocol has been used successfully in our lab to identify the consensus binding sequences for more than 20 additional ZAD family members and has been used by other colleagues in the laboratory to select sequences for other *Drosophila* and mammalian zinc finger transcription factors. Sequences selected in this manner interacted as strongly and specifically as sequences identified in our lab by the traditional all-labeled BSS method. It was also possible to perform selections on nearly three times as many transcription factors/radiolabel batch or to order ~67% less radiolabel for a given set of selections. Our label use efficiency increased by the >50% reduction in rounds of labeled binding. This is because the overall length of time to complete the BSS protocol was also reduced by 50%. This greatly reduced the loss of effective counts to radioactive decay and eliminated complications caused by phosphorylation with partially decayed label. The advantages of reducing the total label required and the time in which the

laboratory must house radioactive isotopes are significant; advantages include decreased costs of label, decreased exposure times for personnel, and fewer survey and storage requirements, as the laboratory can clear of isotope sooner.

ACKNOWLEDGMENTS

Funding assistance from the Department of Biological Sciences, Florida Atlantic University, as a grant-in-aid to K.A. and from NIH to K.A. (Grant 5K01CA95620) to carry out this work is gratefully acknowledged. The authors thank Edna Gamliel for assistance in GST-ZnF protein purification.

REFERENCES

1. Blanton J, Gaszner M, Schedl P. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev* 2003; 17:664–675.
2. Klug A, Rhodes D. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harb Symp Quant Biol* 1987;52: 473–482.
3. Collins T, Stone JR, Williams AJ. All in the family: the BTB/ POZ, KRAB, and SCAN domains. *Mol Cell Biol* 2001;11: 3609–3615.
4. Klug A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem* 2010;79:213–231.
5. Gogos JA, Hsu T, Bolton J, Kafatos FC. Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. *Science* 1992;257:1951–1955.
6. Liu PQ, Rebar EJ, Zhang L, et al. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem* 2010;276:11323–11334.
7. Choo Y, Castellanos A, Garcia-Hernandez B, Sanchez-Garcia I, Klug A. Promoter-specific activation of gene expression directed by bacteriophage-selected zinc fingers. *J Mol Biol* 1997;273:525–532.
8. Fu F, Sander JD, Maeder M, et al. Zinc finger database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res* 2009;37:D279–D283.
9. Cathomen T, Joung JK. Zinc-finger nucleases: the next generation emerges. *Mol Ther* 2008;16:1200–1207.
10. Choo Y, Klug A. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci USA* 1994;91:11163–11167.
11. Wu H, Yang WP, Barbas CF III. Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci USA* 1995;92:344–348.
12. Thiesen HJ, Bach C. Target detection assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res* 1990;18:3203–3209.
13. Swirnoff AH, Milbrandt J. DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol Cell Biol* 1995; 15:2275–2287.
14. Beerli RR, Dreier B, Barbas CF III. Positive and negative regulation of endogenous genes by designed transcription factors. *Proc Natl Acad Sci USA* 2000;97:1495–1500.
15. Payre F, Buono P, Vanzo N, Vincent A. Two types of zinc fingers are required for dimerization of the serendipity delta transcriptional activator. *Mol Cell Biol* 1997;17:3137–3145.
16. Naidu PS, Ludolph DC, To RQ, Hinterberger TJ, Konieczny SF. Myogenin and MEF2 function synergistically to activate the MRF4 promoter during myogenesis. *Mol Cell Biol* 1995;15: 2707–2718.
17. Sekido R, Murai K, Funahashi J, et al. The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation. *Mol Cell Biol* 1994;14:5692–5700.
18. Harms E, Chu T, Henrion G, Strickland S. The only function of Grauzone required for *Drosophila* oocyte meiosis is transcriptional activation of the cortex gene. *Genetics* 2000;155:1831–1839.
19. Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF III. Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 2001; 276:29466–29478.
20. Isalan M, Klug A, Choo Y. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* 1998;37:12026–12033.
21. Greisman HA, Pabo CO. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* 1997;275:657–661.
22. Chung HR, Lohr U, Jackle H. Lineage-specific expansion of the zinc finger associated domain ZAD. *Mol Biol Evol* 2007;24: 1934–1943.
23. Peng H, Zheng L, Lee WH, Rux JJ, Rauscher FJ III. A common DNA-binding site for SZF1 and the BRCA1-associated zinc finger protein, ZBRK1. *Cancer Res* 2002;62:3773–3781.
24. Reiss DJ, Mobley HL. Determination of target sequence bound by PapX, repressor of bacterial motility, in flhD promoter using systematic evolution of ligands by exponential enrichment (SELEX) and high throughput sequencing. *J Biol Chem* 2011; 286:44726–44738.
25. Wang J, Lu J, Gu G, Liu Y. In vitro DNA-binding profile of transcription factors: methods and new insights. *J Endocrinol* 2011;210:15–27.